Balanced Labeled Trees: Density, Complexity and Mechanicity

Nicolas Gast, Bruno Gaujal

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Balanced Labeled Trees: Density, Complexity and Mechanicity

Nicolas Gast — Bruno Gaujal

N° 6240

Juillet 2007

Thème NUM

Rapport de recherche

ISSN 0249-6399

Balanced Labeled Trees: Density, Complexity and Mechanicity

Nicolas Gast , Bruno Gaujal

Thème NUM — Systèmes numériques
Projets MESCAL

Rapport de recherche n° 6240 — Juillet 2007 — 25 pages

Abstract: Sturmian words are very particular infinite words with many equivalent definitions: minimal complexity of aperiodic sequences, balanced sequences and mechanical words. One natural way to generalize the first definition to trees is to consider planar trees of complexity $n+1$ but in doing so, one looses the other two properties.

In this paper we will study non-planar trees. In this case, we show that strongly balanced trees are exactly mechanical trees. Moreover we will see that they are also aperiodic trees of minimal complexity, so that the three equivalent definitions of Sturmian words are almost preserved for non planar trees.

Key-words: Sturmian words, infinite trees, Sturmian trees

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Arbres équilibrés étiquetés : densité, complexité et méchanicité

Résumé : Les mots de Sturm sont des mots infinis qui admettent plusieurs définitions équivalentes: ce sont les mots de complexité minimale, ce sont les mots équilibrés apériodiques et ce sont les mots méchaniques. Une manière naturelle de les généraliser aux arbres (planaires) est d' étendre la notion de complexité. Ce faisant, on perd les deux autres définitions équivalentes.

Une autre manière de généraliser aux arbres binaires infinis est de considérer les arbres non-planaires. Dans ce cas, on montre qu'une généralisation naturelle de la construction mécanique aux arbres permet de montrer que les arbres mécaniques sont aussi les arbres fortement équilibrés et qu'ils sont de complexité minimale, ce qui permet de retrouver, dans une certain mesure, les équivalences existantes pour les mots de Sturm.

Mots-clés : Mots de Sturm, arbres infinis, arbres de Sturm

1 Introduction

Sturmian words are infinite words over a binary alphabet, say $\{0,1\}$, that have exactly $n+1$ factors of length n. They also admit other equivalent definitions (see [4], for a rather exhaustive presentation of Sturmian words).

Definition 1.1. *A word $w \in \{0,1\}^{\mathbb{N}}$ is Sturmian word if it verifies of the three equivalent properties.*
(i) For all $n \geq 0$: w has exactly $n+1$ factor of length n.
(ii) w is balanced and aperiodic: if x and y are two factors of length n and if we denote by $|x|_1$ the number of 1 in x, then $\Big||x|_1 - |y|_1\Big| \leq 1$.
(iii) w is a mechanical word with an irrational slope: there exist $\alpha \in (0;1) \setminus \mathbb{Q}$ and $\phi \in [0;1[$ such that: for all i, $w_i = \lfloor (i+1)\alpha + \phi \rfloor - \lfloor i\alpha + \phi \rfloor$ *or* $w_i = \lceil (i+1)\alpha + \phi \rceil - \lceil i\alpha + \phi \rceil$.

In [2], Berstel and al. generalized this notion to Sturmian trees which are planar binary trees with complexity $n+1$. These trees are irrational planar trees with minimal complexity. However, unlike in the case of words, it seems that there is no simple equivalent definition. In particular, there is no link with the balance property or with the mechanical construction for these trees.

Here, we will focus on *non-planar* balanced tree, showing that they provide several existence and equivalent definitions in the flavor of Sturmian words. In particular, we show that trees are strongly balanced if and only if they are mechanical. Also mechanical trees are shown to have minimal complexity.

1.1 Definitions

Throughout this paper, we will focus on trees which are rooted, binary, infinite, labelled by $\{0,1\}$ and non planar (*i.e.* there is no distinction between the "left" or "right" sub-tree). More precisely:

Definition 1.2 (Infinite tree)**.** *A tree is a triplet $(\mathfrak{E}, \mathbf{P}, f)$ where:*

1. *$\mathfrak{E} \subset \mathbb{N}$ is the set of nodes.*
2. *$\mathbf{P} : \mathfrak{E} \to \mathfrak{E}$ has the following properties:*
   - *There exists a unique r such that $\mathbf{P}(r) = r$ (r is the root of the tree)*
   - $\forall n \neq r : \ \mathbf{P}(n) < n$
   - $\forall n \neq r : \mathrm{card}(\{x | \mathbf{P}(x) = n\}) = 2$ *(the tree is binary).*
   - $\mathrm{card}(\{x | \mathbf{P}(x) = r\}) = 3$.
3. *$f : \mathfrak{E} \to \{0;1\}$. $f(n)$ is the label of n.*

Definition 1.3 (Rooted sub-tree, non-rooted sub-tree)**.** *Let $\mathcal{A} = (\mathfrak{E}, \mathbf{P}, f)$ be a tree.*

The sub-tree of root $n \in \mathfrak{E}$ and height $k \in \mathbb{N} \cup \{+\infty\}$, $\mathcal{A}_{[n,k]} = (\mathfrak{E}_{[n,k]}, \mathbf{P}_{[n,k]}, f_{[n,k]})$ is the tree defined by:

- $\mathfrak{E}_{[n,k]} = \big\{x | \exists q < k \text{ such that } \mathbf{P}^q(x) = n\big\}$
- *$\mathbf{P}_{[n,k]}$ is the restriction of $\mathbf{P}$ to $\mathfrak{E}_{[n,k]}$ with $\mathbf{P}_{[n,k]}(n) = n$.*

- *$f_{[n,k]}$ is the restriction of f to $\mathfrak{E}_{[n,k]}$.*

A non-rooted sub-tree of height k and width 2^q, $(q < k)$ is the restriction of a tree to the subset $\mathfrak{E}_{[n,k+q]} \setminus \mathfrak{E}_{[n,q]}$.

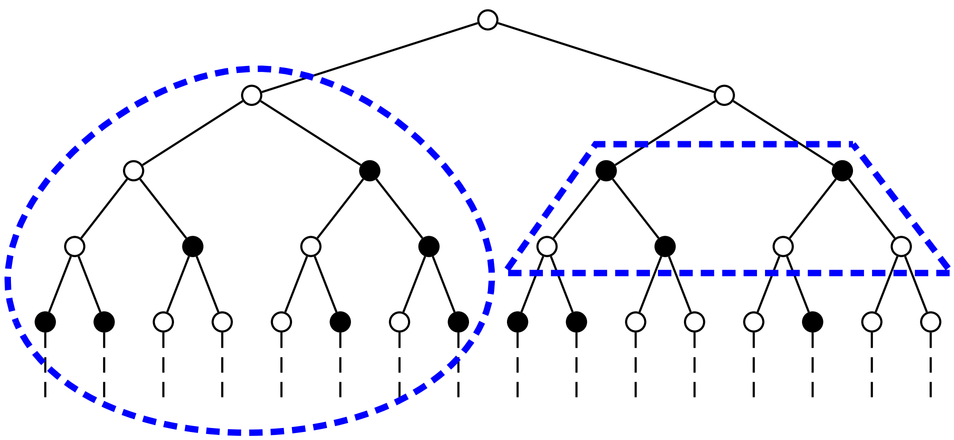


Figure 1: The left sub-tree is a rooted sub-tree of height 4, the right sub-tree is a non-rooted sub-tree of height 2 and width 2.

In the rest of the document, we will call tree either an infinite tree or a sub-tree of an infinite tree. We will also set the height of an infinite tree $\mathcal{B}$ to be ∞ and to be k if $\mathcal{B} = \mathcal{A}_{[n,k]}$ for some tree $\mathcal{A}$.

The number of node of a tree $\mathcal{B}$ is card($\mathbb{N}$) if $\mathcal{B}$ infinite and $2^k - 1$ if $\mathcal{B} = \mathcal{A}_{[n,k]}$.

Canonical representation

Two trees $(\mathfrak{E}, \mathbf{P}, f)$, $(\mathfrak{E}', \mathbf{P}', f')$ are equivalent if there exists a bijection $b : \mathfrak{E} \mapsto \mathfrak{E}'$ such that: $b(r) = r'$, $f(n) = f'(b(n))$ and $b(\mathbf{P}(n)) = \mathbf{P}'(b(n))$

Let $\mathcal{A}$ be a tree, we can choose a canonical representation of $\mathcal{A}$ by a word on $\{0, 1\}$: if we call $\{\mathcal{B}\}$ the set of all the trees equivalent to $\mathcal{A}$ such that $\forall k, \forall 0 \leq i \leq k-1$: $\mathbf{P}_{\mathcal{B}}(2^{k+1}+2i) = 2^k+i$ and $\mathbf{P}_{\mathcal{B}}(2^{k+1}+2i+1) = 2^k+i$, we can define for each $\mathcal{B}$ a word $u_{\mathcal{B}}$ by $u_i = f_{\mathcal{B}}(i)$. Among all the words $u_{\mathcal{B}}$, the word $w_{\mathcal{A}}$ associated with $\mathcal{A}$ is the minimal one (using the lexical order).

Figure 2 gives an example of a tree and the canonical word associated.

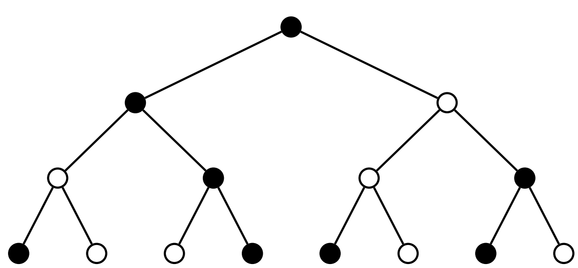


Figure 2: Example of a tree of height 4, whose canonical representation is 101010101010101.

Density

Let $\mathcal{A}$ be an infinite tree, n a node and $k \geq 0$. We define $h(\mathcal{A}_{[n,k]})$ to be the number of nodes labeled by 1 is the sub-tree of root n and height k:

$$h(\mathcal{A}_{[n,k]}) = \sum_{i \in \mathfrak{E}_{[n,k]}} f(i).$$

The density of a sub-tree $\Pi(\mathcal{A}_{[n,k]})$ by the number of 1 on its total number of nodes:

$$\Pi(\mathcal{A}_{[n,k]}) = \frac{h(\mathcal{A}_{[n,k]})}{2^k - 1}.$$

We want to call density of an infinite tree $\mathcal{A}$, the average number of 1 in the tree. Actually, this quantity is the limit of the density of sub-trees of height k (it may or may not exist). We say that the density of a tree is $\alpha \in \mathbb{R}$ *if and only if*

$$\forall n, \quad \lim_{k \to \infty} \Pi(\mathcal{A}_{[n,k]}) = \alpha.$$

1.2 Rational trees: complexity and density

For words, there is a simple definition of what a periodic word is: it is a finite factor that *repeats* itself. The case of trees is a little more complicated. The property captured in the definition below is that a finite number of finite patterns will define the infinite tree.

Definition 1.4. *Let $\mathcal{A}$ be a infinite tree. We call $S_n(\mathcal{A})$ the set of the equivalence class of sub-trees of $\mathcal{A}$ of height n.*

We denote by $P(\mathcal{A}, n)$ the number of this class:

$$P(\mathcal{A}, n) = \mathrm{card}(S_n(\mathcal{A})).$$

A tree has always a finite number of sub-trees of height k (bounded by $2^{2^k - 1}$). Let $n \geq 0$, we call *factor graph of order n* of $\mathcal{A}$ the graph $G_n = (S_n, E_n)$ defined by:

- S_n is the set of sub-trees of $\mathcal{A}$ defined above.
- a tuple (F, C_1, C_2) belongs to $E_n \subset S_n \times (S_n \times S_n)$ *iff* there exists three nodes f, c_1, c_2 such that c_1 and c_2 are the two children of f and f, c_1, c_2 are roots of sub-trees respectively equivalent to F, C_1, C_2. In that case, we say that there is an edge from F to $\{C_1, C_2\}$.

An example of such a graph is shown figure 3

Let us consider an infinite tree and let u be a node. The signification of this graph is that if the sub-tree of size n corresponding to u is F, its two children will be in the set $\{\{C_1, C_2\}/(F, C_1, C_2) \in \mathfrak{E}_n\}$. For example if the graph has exactly one outgoing edge for each vertices F, the tree is fixed by the graph and its first sub-tree.

Definition 1.5 (Rational tree)**.** *Let $\mathcal{A}$ be an infinite tree labeled with an alphabet with k elements. We said that $\mathcal{A}$ is rational if it satisfy one of the three equivalent properties:*

(i) $\{P(\mathcal{A}, n)\}_n$ *is bounded*

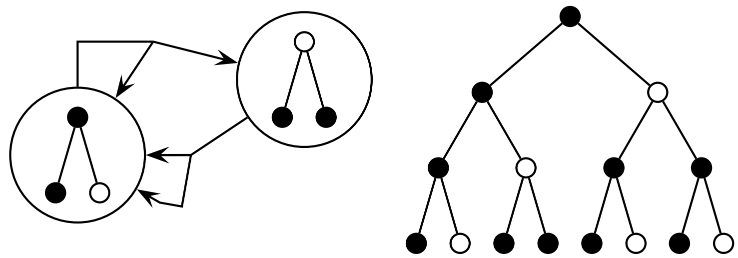


Figure 3: A rational tree and its factor graph.

(ii) $P(\mathcal{A}, n) = P(\mathcal{A}, n+1)$ *for one* n.

(iii) $P(\mathcal{A}, n) < n + k - 1$ *for one* n *where* k *is the number of different letters appearing in* $\mathcal{A}$.

Proof. (i) implies (iii) is clear.

(iii) implies (ii): if $P(\mathcal{A}, i) < P(\mathcal{A}, i+1)$ for $i < n$, then $P(\mathcal{A}, n) \geq n - 1 + P(\mathcal{A}, 0) = n - 1 + k$, which contradicts (iii).

(ii) implies (i): for all n, we can define the factor graph of $\mathcal{A}$. As each node of this graph is a factor of the tree there is at least one edge going out of each node. Moreover the number of outgoing edges is the number of factor of length $n+1$ so it is $P(\mathcal{A}, n+1) = P(\mathcal{A}, n)$. Thus there is one and only one edge going out of each node. That means that if we start from a vertex of the graph, we will follow a deterministic path so there is exactly $P(\mathcal{A}, n)$ factors of length $k \leq n$ □

Proposition 1.6. *Let* $\mathcal{A}$ *be a rational tree. If* $\mathcal{A}$ *has a density* α*, then* α *is rational.*

Proof. (sketch) As $\mathcal{A}$ is rational, there exists k such that the tree is completely defined by its sub-trees $(A_1, \ldots, A_k)$ of height k. As the tree is rational, each sub-tree A_i has two children (A_{i_1}, A_{i_2}). We consider a Markov chain (X_n) on the set $\{A_1, \ldots, A_k\}$:

$$\mathbf{P}(X_{n+1} = A_{i_1} | X_n = A_i) = \mathbf{P}(X_{n+1} = A_{i_2} | X_n = A_i) = \frac{1}{2}.$$

A sequence $X_0, \ldots, X_n, \ldots$ defines a unique path in the tree corresponding to a word w where w_i is the label of the root of A_i. We call density of this path the density of the word w if it exists.

If the Markov chain is irreducible and aperiodic, there exists α such that all paths w have almost surely a density α corresponding to the stationary distribution of the chain, which is a solution of a rational linear system and is therefore rational.

let us pick a sub-tree according to Π. Let w be a random path of length n beginning at the root of this tree. It is also distributed according to the stationary distribution. Now let us pick two random paths of length $n-1$ in the rest of the sub-tree (disjoint from the previous path),...

All of these path are distributed according the stationary distribution. Let C_i be the distribution of the number of 1 in a path of length $n-i$. If the number of a tree of length n is distributed as B_n, we have the relation:

$$\begin{aligned} h(B_n) &\sim C_n + B_{n-1} + B_{n-2} + B_{n-3} + \cdots + B_1 \\ &\sim C_n + C_{n-1} + 2C_{n-2} + 2^2 C_{n-3} + \cdots + 2^{n-2} C_1, \end{aligned} \tag{1}$$

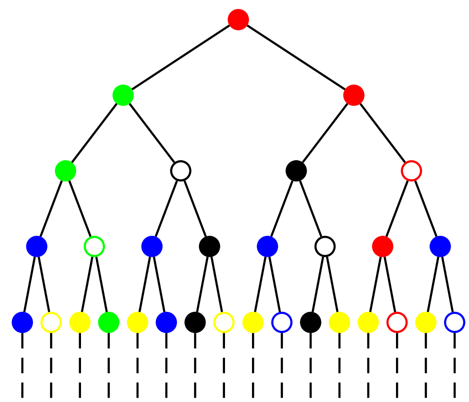


Figure 4: Example of decomposition of a tree of height 5 in paths of length 5,4,3,2 and 1. Each color corresponding to a path.

B_n is a sum of C_n thus in average, the density of A_n is α.

The number of 1 of the sub-tree of height $2n$ is the number of 1 in the sub-tree A_n of height n plus 2^n times the number of 1 in the sub-trees $(a_1, \ldots, a_{2^n})$ of height n that are children of A_n. Thus the number of 1 in A_n is negligible compared to the number in $(a_1, \ldots, a_{2^n})$.

If the Markov chain is aperiodic, the a_i's are distributed according to the stationary law and then the number of 1 is α. If the chain is periodic the chain might have a density (and it is α) but also might not have a density, in that case, we say that α is the periodic-density of the tree. See figure 6 for two examples of rational trees that do or do not have density.

If the chain is reducible, then the analysis can be carried using the "one step" technique for transient Markov chains

If the chain is reducible, the set of its states can be divided in subsets $S_1, \ldots, S_m$ on which the chain is irreducible plus a set $\mathcal{O}$ of the other states that belong to no S_i. Let α_i be the density of the paths in the subset S_i.

If we start inside an S_i, we know that we will stay forever in this S_i. What we want to show is that if we start outside the set of the S_i, we have for all i a rational probability of ending in S_i.

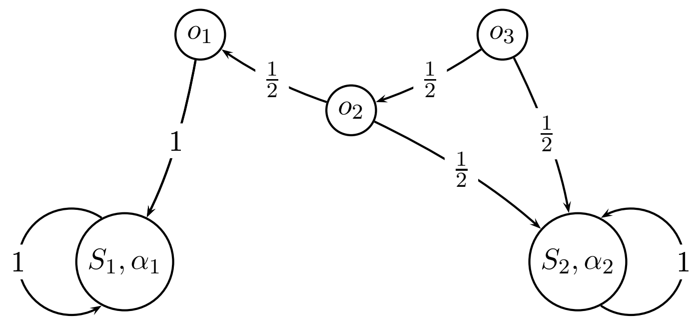
If there is l states in $\mathcal{O}$, we define a matrix X of size $l \times l$: X_{ij} is the probability of transition from i to j (which equals 0, 1 or 1/2). For each irreducible subset S_i, let us define a_i a variable and let us define the vector (of size l) R where for all $o \in \mathcal{O}$:

$$R_o = \sum_i \mathbf{P}(X_{n+1} \in S_i | X_n = o) a_i.$$

Each R_o is a linear combination on the set $\{a_i\}$. Let us define a vector $P = \sum_i \mathbf{P}(\exists n / X_n \in S_i | X_0 = o) a_i$ where $\mathbf{P}(\exists n / X_n \in S_i | X_0 = o)$ the probability of ending in a S_i starting in o, it satisfies the equality

$$P = XP + R.$$

In order to prove our result, we want to prove that this system has a unique solution and that this solution is rational.



The corresponding matrices X,R and P are:

$$X = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad R = \begin{bmatrix} a_1 \\ \frac{1}{2}a_2 \\ \frac{1}{2}a_2 \end{bmatrix} \qquad P = \begin{bmatrix} a_1 \\ \frac{1}{2}(a_1 + a_2) \\ \frac{1}{4}(a_1 + 3a_2) \end{bmatrix}.$$

Figure 5: Example of a reducible Markov chain. S_i and S_2 are the two irreducible subset. We assume that each S_i is periodic and has density α_i. Then a tree beginning with o_1 will have density α_1, with o_2 $\frac{\alpha_1+\alpha_2}{2}$ and with o_3: $\frac{\alpha_1+3\alpha_2}{4}$.

As no element of $\mathcal{O}$ belongs to an irreducible subset, for all subsets $O \subset \mathcal{O}$, there exists at least one state in O such that the probability of leaving O is greater than 0: let us call it j_0. We have $\sum_{i\in O} X_{j_O i} < 1$ (and of course for all $i \in O$: $\sum_{i\in O} X_{ij} \leq 1$).

If there exists a vector x such that $X.x = x$, let us call O the subset $O = \{i/x_i \neq 0\}$. Then there exists j_O such that $\sum_{j\in O} X_{ij} < 1$ and then $\sum_i x_i = \sum_i \sum_j X_{ij}x_j \leq \sum_i \sum_j X_{ij} \sum_j x_j < \sum_j x_j$ which contradicts $X.x = x$. Consequently, the matrix $\mathbf{Id} - X$ is irreducible.

The system $(\mathbf{Id} - X)P = R$ has a unique solution since $\mathbf{Id} - X$ is irreducible and for all o, there exists P_{oi} such that:

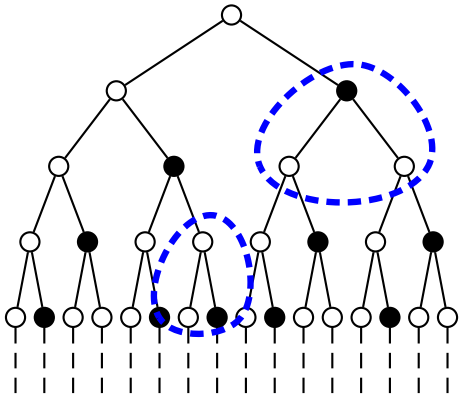
$$R_o = \sum_i P_{oi} a_i,$$

P_{oi} is the probability that starting from the state o, the Markov chain will end in the subset S_i. Moreover all P_{oi} are rational since all numbers X_{oj} and R_o are rational.

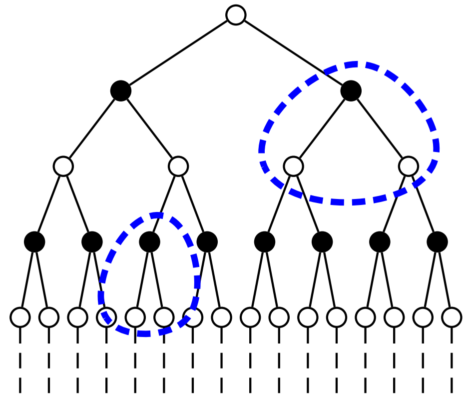
Thus if the root of the tree is r, as k tends to infinity, a sub-tree has in average a density α where

$$\alpha = \sum_i P_{oi} \alpha_i.$$

Thus the periodic density of the tree is α. That is why if the tree has a density, it will be $\alpha \in \mathbb{Q}$. □



The tree is rational since it has two different sub-trees of height 2 (surrounded by blue in the figure). The corresponding Markov chain is aperiodic and then it has a density.



This one has two sub-trees of height 2. The corresponding Markov chain is periodic.

Figure 6: This figure represents two rational trees. The Markov chain corresponding to the left one is aperiodic and thus it has a density: 1/3. The Markov chain corresponding to the left one is periodic and this tree does not have a density but its periodic-density is $\frac{1}{2}$.

2 Balanced trees

An infinite word w is said to be balanced if for all $n > 0$, two subsequences of w of length n will have almost the same density of 1. More precisely, if we call $h(x)$ the number of 1 in the word x, w is balanced *iff* for all $n > 0$ and for all x, y of length n:

$$|h(x) - h(y)| \leq 1.$$

A balanced word has many properties. In particular, a balanced word always admits a density (which means $\lim_{n\to\infty} \frac{h(w_1 \dots w_n)}{n}$ exists) and an aperiodic balanced word is Sturmian.

In this document, we will study two generalizations to trees of this definition. The first one and probably the most natural one, is what we call a *balanced tree*: a tree is said to be balanced if all of its rooted sub-trees of height k have almost the same number of 1:

Definition 2.1 (Balanced tree)**.** *A infinite tree $\mathcal{A}$ is* balanced *if for all node n and for all $k \geq 0$:*

$$|h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n',k]})| \leq 1.$$

Let us also introduce a stronger definition: a tree is *strongly balanced* if all of its non-rooted sub-tree of height k and wide w have (almost) the same number of 1:

Definition 2.2 (Strongly balanced tree)**.** *A tree $\mathcal{A}$ is said to be* strongly balanced *if for all $k, q \geq 0$ and for all nodes n, n':*

$$\Big| |h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n,q]})| - |h(\mathcal{A}_{[n',k]}) - h(\mathcal{A}_{[n',q]})| \Big| \leq 1.$$

This latest definition is clearly stronger than the standard one since taking $q = 0$ implies $\left|h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n',k]})\right| \leq 1$. We will see in the next part that this definition is strictly stronger since there exist trees that are balanced but not strongly balanced.

Although the definition of a balanced tree is weaker and seams more natural for a generalization from words, we will see that strongly balanced tree have almost the same properties than its counterpart on words.

2.1 Density of a balanced tree

In this part, $\mathcal{A}$ denotes a balanced tree. As for the case of balanced sequences, we can define the density of a balanced tree.

For all n, $|h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n',k]}| \leq 1$ thus for all k we can define m_k which as the minimum of $h(\mathcal{A}_{[n,k]})$ over all n.

$$m_k \leq h(\mathcal{A}_{[n,k]}) \leq m_k + 1.$$

A sub-tree of height $q+k$ is composed of a tree of size q which has between m_q and m_q+1 ones with 2^q children of size k which have between m_k and m_k+1 ones. Thus:

$$2^q m_k + m_q \leq m_{q+k} \leq 2^q(m_k+1) + m_q + 1.$$

Proposition 2.3 (Density of balanced tree)**.** *Let $\mathcal{A}$ be a balanced tree.*

There exists a unique α, called the density *of $\mathcal{A}$ such that for all n:*

$$\lim_{k\to\infty} \Pi(\mathcal{A}_{[n,k]}) = \alpha.$$

Moreover

$$\forall n,k : |h(\mathcal{A}_{[n,k]}) - \lfloor (2^k-1)\alpha \rfloor| \leq 1.$$

Before starting with the proof, we can remark that the tree represented by figure 2: 101010101010101... is balanced with density $\frac{1}{2}$.

Proof. The proof simply uses the definition of m_q to bound $\Pi(\mathcal{A}_{[n,k+q]}) - \Pi(\mathcal{A}_{[n,k]})$ on both sides.

1. Let n be a node, $k, q \geq 0$.

Using the definition of m_k, we can bound the value $\Pi(\mathcal{A}_{[n,k+q]}) - \Pi(\mathcal{A}_{[n,k]})$:

$$\frac{m_{q+k}}{2^{q+k}-1} - \frac{m_k+1}{2^k-1} \leq \Pi(\mathcal{A}_{[n,k+q]}) - \Pi(\mathcal{A}_{[n,k]}) \leq \frac{m_{q+k}+1}{2^{q+k-1}-1} - \frac{m_k}{2^k-1}.$$

Considering the fact that $2^q m_k + m_q \leq m_{q+k} \leq 2^q(m_k+1) + m_q + 1$, we have

$$\frac{2^q m_k + m_q}{2^{q+k}-1} - \frac{m_k+1}{2^k-1} \leq \Pi(\mathcal{A}_{[n,k+q]}) - \Pi(\mathcal{A}_{[n,k]}) \leq \frac{2^q(m_k+1) + m_q + 1}{2^{q+k-1}-1} - \frac{m_k}{2^k-1}.$$

The absolute value of the right term is
(regardless of q) clearly smaller than ϵ as soon as k is big enough since $\frac{m_k}{2^k} \leq 1$.

Thus $(\Pi(\mathcal{A}_{[n,k]}))_k$ is a Cauchy sequence and there exists α such that $\lim_{k\to\infty}(\Pi(\mathcal{A}_{[n,k]}))_k = \alpha$.

This limit does not dependant on n since the tree is balanced.

2. Lets now prove that $|h(\mathcal{A}_{[n,k]}) - \lfloor (2^k - 1)\alpha \rfloor| \leq 1$:
We have the equality

$$2^q m_k + m_q \leq m_{q+k} \leq 2^q(m_k + 1) + m_q + 1.$$

Dividing by $2^{k+q} - 1$ and looking the limit towards ∞, we have:

$$\frac{m_k + \alpha}{2^k} \leq \alpha \leq \frac{m_k + 1 + \alpha}{2^k}.$$

Thus

$$(2^k - 1)\alpha - 1 \leq m_k \leq (2^k - 1)\alpha.$$

□

If $\mathcal{A}$ is a strongly balanced tree, there is a very similar formula linking its density and the number of 1 in one of its sub-tree: let us assume that $k > q > 0$, then for all node n, we have:

$$\left|h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n,q]}) - (2^k - 2^q)\alpha\right| \leq 1. \tag{2}$$

The remarkable property about density it that it does not depends on the sub-tree chosen: there exists α such that the density of any sequence of sub-trees of increasing height will converge to α. Naturally, many trees do not have a density. For example the tree which line i is labeled by i mod 2 does not have a density. See figure 6 for examples of trees having or not a density.

Also, deciding if a finite tree is balanced can be done in linear time in the number of nodes of the tree.

```
is_balanced (A):
  k <- height(A)
  for i=1 to k
    min(i) = +infinity
    max(i) =-infinity

  for i=1 to k
    visiting the tree by a level order traversal, compute the number of 1
    in all sub-tress which leaves are at height i:

    for each tree that we visit, let us call j its height j and h its
    number of nodes labelled by 1:
      if min(j) > h
        min(j) = h
      if max(j) < h
        max(j) = h
      if max(j) - min(j) >= 2
        return false

  return true
```

If the height of the tree is k, visiting the sub-tree by a level order down to the height i can be done in time $O(2^{k-i})$. Thus the complexity of the algorithm is $\sum_i O(2^{k-i}) = O(2^k)$ which means that it runs linear time in the number of nodes. You can see an example of an execution of the algorithm on figure 7. Applying the same ideas, Similarly, testing if a tree of height k is strongly balanced can be done in time $O(2^k)$ and size $O(k^2)$.

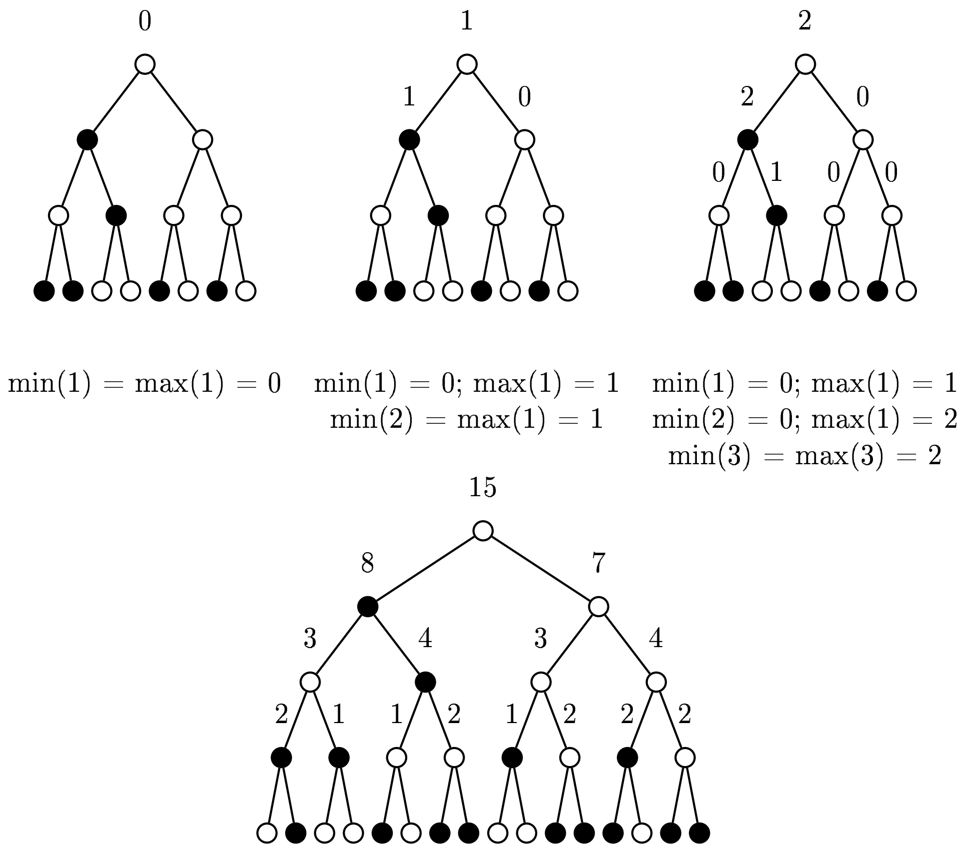


Figure 7: Example of the execution of the algorithm that decides if a tree is balanced. In this particular case, the tree is not balanced since max(2)-min(2) = 2.

3 Mechanical trees

Let us recall the definition of a mechanical words: w is a mechanical word with slope α if there exist $\phi \in [0, 1)$ such that:

$$\text{for all } i : w_i = \lfloor (i+1)\alpha + \phi \rfloor - \lfloor i\alpha + \phi \rfloor.$$

Sturmian words can be also defined as aperiodic balanced words or as mechanical words of irrational slope. In this part, we will see that we have the same equivalence properties between strongly balanced trees and mechanical trees.

Definition 3.1 (Mechanical tree)**.** *A tree $\mathcal{A}$ is said to be mechanical of density α if for all node n:*

$$\exists \phi_n \in [0,1[\ \forall k:\ h(\mathcal{A}_{[n,k]}) = \lfloor (2^k-1)\alpha + \phi_n \rfloor \text{ or } \exists \phi_n \in [0,1[\ \forall k:\ h(\mathcal{A}_{[n,k]}) = \lceil (2^k-1)\alpha - \phi_n \rceil.$$

If the node n of an mechanical tree verifies the relation $\forall k:\ h(\mathcal{A}_{[n,k]}) = \lfloor (2^k-1)\alpha + \phi \rfloor$, this node is said inferior of phase ϕ. In fact it is an abuse of notation to say so because there could exist ϕ_1 and ϕ_2 such that for all k: $h(\mathcal{A}_{[n,k]}) = \lfloor (2^k-1)\alpha + \phi_1 \rfloor = \lfloor (2^k-1)\alpha + \phi_2 \rfloor$. However to simplify the notation, when the phase of a sub-tree is said to be ϕ, it could be any ϕ that works.

Let us call $\text{frac}(x)$ the fractional part of a real number x and let us look at the sequence $\left(\text{frac}(2^k\alpha - \alpha + \phi)\right)_k$. If this sequence can be arbitrarily close to 0, this means that if $\psi > \phi$, there exists k such that $\lfloor (2^k-1)\alpha + \psi \rfloor > \lfloor (2^k-1)\alpha + \phi \rfloor$. On the other hand, if this sequence can be arbitrarily close to 1, if $\psi < \phi$, then there exists k such that $\lfloor (2^k-1)\alpha + \psi \rfloor < \lfloor (2^k-1)\alpha + \phi \rfloor$.

Therefore, a phase ϕ is unique *iff* $\left(\text{frac}((2^k-1)\alpha + \phi)\right)_k$ is arbitrarily close to 0 and 1.

Now let us call $\alpha_1, \alpha_2, \dots, \alpha_n, \dots \in \{0;1\}^{\mathbb{N}}$ the sequence of digits of α in base 2 and $x_1, \dots, x_n, \dots$ the sequence of digits of $\alpha - \phi$ in base 2. A multiplication by 2^k corresponds to a shift of the digits by k. If $x_1, \dots, x_k$ does not end with neither an infinite number of 0 nor an infinite sequence of 1[1], $\left(\text{frac}((2^k-1)\alpha + \phi)\right)_k$ is arbitrarily close to 0 or 1 means that for all n, there exists k such that $\alpha_k, \dots, \alpha_{k+n-1} = x_1, \dots, x_n$. More precisely, if for all n, there exists $m > n$ and $k > 0$ such that $x_m = 0$ (resp. 1) and $\alpha_k, \dots, \alpha_{k+m-1} = x_1, \dots, x_{m-1}, 1$ (resp. $x_1, \dots, x_{m-1}, 0$), then $\left(\text{frac}((2^k-1)\alpha + \phi)\right)_k$ is arbitrarily close to 0.

Thus we have to distinguish three cases:

- If α is a number such that all finite binary sequences appear in the binary expansion of α, then for all phase ϕ, then ϕ is unique. In particular, all *normal numbers* verify this property and we know that almost every number in $[0,1]$ is normal, see [3] (a number is normal if all sequences of length k appear with rate 2^{-k} in the binary expansion of α).
- If $\alpha \in \mathbb{Q}$, then the sequence $\text{frac}((2^k-1)\alpha + \phi))$ is periodic and there are no phase ϕ such that ϕ is unique.
- If α is neither rational nor has the property that all binary sequences appear in α, then some ϕ can be unique and some others may not. For example, if α is the number:

  $$\alpha = 0.10110011100011110000111110000\dots,$$

  then if $\text{frac}(\alpha - \phi) = 0$, ϕ is unique, while ϕ_1 and ϕ_2 such that $\text{frac}(\alpha - \phi_1) = 0.10100$ and $\text{frac}(\alpha - \phi_2) = 0.1010$ are equivalent.

If one fixes the density α and the phase ϕ_n of each node, there is at most one matching mechanical tree. The first result of this part is to prove that if we choose the density α and the phase ϕ_0 of the root, there exists one and only one mechanical tree.

Moreover in our definition, we allow some nodes to be *inferior* (which means that it is of the form $\forall k:\ h(\mathcal{A}_{[n,k]}) = \lfloor (2^k-1)\alpha + \phi \rfloor$) and some nodes to be *superior* (which means

[1] The case where x ends with an infinite number of 0 (or 1) is quite similar: if $x_1, x_2, \dots, = x_1, x_l, 1, 0, 0, 0, \dots$, that means that $\left(\text{frac}((2^k-1)\alpha + \phi)\right)_k$ is arbitrarily close to 1 if for all m, there exists k such that $\alpha_k, \dots, \alpha_{k+n-1} = x_1, \dots, x_l, 0, 0, 0, \dots, 0$

$h(\mathcal{A}_{[n,k]}) = \lceil (2^k-1)\alpha+\phi \rceil)$. We will see that if a node if inferior, its two children will be also inferior.

Proposition 3.2. *Let $\alpha \in [0,1], \phi \in [0,1[$.*
There exists a unique mechanical tree $\mathcal{A}$ of density α and initial phase ϕ_0. which means:

$$\forall k: \ h(\mathcal{A}_{[0,k]}) = \lfloor (2^k-1)\alpha+\phi \rfloor.$$

Proof. (sketch) It is based on some relations on rounding functions that will prove that if a node has a phase ϕ, the phases of its two children are fixed. □

Lemma 3.3. *Let $x \in \mathbb{R}$, then*

$$\left\lceil \frac{\lceil x \rceil}{2} \right\rceil = \left\lceil \frac{x}{2} \right\rceil \tag{3}$$

$$\left\lceil \frac{\lfloor x \rfloor}{2} \right\rceil = \left\lfloor \frac{x+1}{2} \right\rfloor \tag{4}$$

$$\left\lfloor \frac{\lfloor x \rfloor}{2} \right\rfloor = \left\lfloor \frac{x}{2} \right\rfloor \tag{5}$$

$$\left\lfloor \frac{\lceil x \rceil}{2} \right\rfloor = \left\lceil \frac{x-1}{2} \right\rceil. \tag{6}$$

Proof. Let $p = \lceil x \rceil$ and $\theta = x - p \in [0;1[$. If $x = p$, 3 is clearly true. Otherwise:

$$\begin{aligned} \left\lceil \frac{\lceil x \rceil}{2} \right\rceil &= \left\lceil \frac{p+1}{2} \right\rceil \\ &= \begin{cases} \frac{p+1}{2} \text{ if p is odd} \\ \frac{p}{2}+1 \text{ if p is even.} \end{cases} \end{aligned}$$

$$\begin{aligned} \left\lceil \frac{x}{2} \right\rceil &= \left\lceil \frac{p+\theta}{2} \right\rceil \\ &= \begin{cases} \left\lceil \frac{p+1}{2} + \frac{\theta-1}{2} \right\rceil = \frac{p+1}{2} \text{ if p is odd} \\ \left\lceil \frac{p+2}{2} + \frac{\theta-1}{2} \right\rceil = \frac{p}{2}+1 \text{ if p is even.} \end{cases} \end{aligned}$$

To show 4, let $p = \lceil x \rceil$ and $\theta = x - p \in [0; 1[$. If $\theta \neq 0$:

$$\begin{aligned}
\left\lceil \frac{\lfloor x \rfloor}{2} \right\rceil &= \left\lceil \frac{\lceil x-1 \rceil}{2} \right\rceil \\
&= \left\lceil \frac{x-1}{2} \right\rceil \\
&= \left\lfloor \frac{x-1}{2} + 1 \right\rfloor \\
&= \left\lfloor \frac{x+1}{2} \right\rfloor .
\end{aligned}$$

If $\theta = 0$, let $x = 2q$.

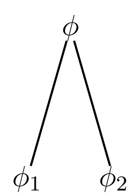
$$\left\lceil \frac{\lfloor x \rfloor}{2} \right\rceil = q,$$

$$\left\lfloor \frac{x+1}{2} \right\rfloor = \left\lfloor q + \frac{1}{2} \right\rfloor = q.$$

The same proof holds for the other inequalities. ☐

Proof. **Uniqueness.** Let $\mathcal{A}$ be a mechanical trees of density α.

The label of the root is $\lfloor \alpha + \phi \rfloor$ so it is completely determined by α and ϕ. As $\mathcal{A}$ is a mechanical tree, its two children (noted $\mathcal{A}_1$ and $\mathcal{A}_2$) are also mechanical trees. We call ϕ_1 and ϕ_2 their phase and since the tree is non-planar, without loose of generality, we assume that $\phi_1 \leq \phi_2$. Moreover we assume that the root is of the form $h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi \rfloor$ (otherwise, we just replace $\lfloor$ by $\lceil$).



Let us compute the numbers ϕ_1 and ϕ_2 by computing $h(\mathcal{A}_{[r,k+1]})$ from two different ways:

$$\forall k : \lfloor (2^{k+1} - 1)\alpha + \phi \rfloor = \lfloor (2^k - 1)\alpha + \phi_1 \rfloor + \lfloor (2^k - 1)\alpha + \phi_2 \rfloor + \lfloor \alpha + \phi \rfloor .$$

If the two children are inferior, as $0 \leq \phi_1 \leq \phi_2 \leq 1$, we got $0 \leq \phi_2 - \phi_1 \leq 1$ and then there exists $c = 0$ or 1 such that:

$$\lfloor (2^k - 1)\alpha + \phi_1 \rfloor \leq \lfloor (2^k - 1 + \phi_2) \rfloor + c.$$

If one child is inferior and the other one is superior, we remark that this number c still exists and it does not change the end of the proof. In fact if the root is inferior, its two children can be expressed as inferior children too.

Thus:

$$\left\lfloor (2^k - 1)\alpha + \phi_1 \right\rfloor = \left\lfloor \frac{\lfloor (2^{k+1} - 1)\alpha + \phi \rfloor - \lfloor \alpha + \phi \rfloor}{2} \right\rfloor.$$

So the only question remaining is whether or not $\left\lfloor \frac{\lfloor (2^{k+1}-1)\alpha+\phi \rfloor - \lfloor \alpha+\phi \rfloor}{2} \right\rfloor$ can be written as $\left\lfloor (21^k - 1)\alpha + \phi_1 \right\rfloor$ and $\left\lceil \frac{\lfloor (2^{k+1}-1)\alpha+\phi \rfloor - \lfloor \alpha+\phi \rfloor}{2} \right\rceil$ can be writ-en as $\left\lfloor (2^k - 1)\alpha + \phi_2 \right\rfloor$.

By lemma 3.3, we can compute the following tabular that shows us different values for ϕ_1 and ϕ_2 depending on the initial ϕ:

$$\phi$$

$$\frac{\alpha+\phi}{2} - \lfloor \alpha + \phi \rfloor \qquad \frac{\alpha+\phi+1}{2} - \lfloor \alpha + \phi \rfloor$$

	$\lfloor \alpha + \phi \rfloor < 1$	$\lfloor \alpha + \phi \rfloor \geq 1,$
ϕ_1	$\frac{\alpha+\phi}{2}$	$\frac{\alpha+\phi+1}{2},$
ϕ_2	$\frac{\alpha+\phi-1}{2}$	$\frac{\alpha+\phi}{2}.$

Remark that if the root is inferior its two children can be expressed as inferior nodes with phase equals to the values in the previous tabular. That means that if a children is inferior and one of its child superior, it can be expressed as a inferior node too.

Le us focus on the second column (the rest is similar) $\lfloor \alpha + \phi \rfloor < 1$

$$\left\lceil \frac{\lfloor (2^{k+1} - 1)\alpha + \phi \rfloor - \lfloor \alpha + \phi \rfloor}{2} \right\rceil = \left\lceil \frac{\lfloor (2^{k+1} - 1)\alpha + \phi \rfloor}{2} \right\rceil \tag{7}$$

$$= \left\lceil \frac{\lceil (2^{k+1} - 1)\alpha + \phi \rceil}{2} - 1 \right\rceil \tag{8}$$

$$= \left\lceil \frac{(2^{k+1} - 1)\alpha + \phi - 1}{2} \right\rceil \tag{9}$$

$$= \left\lfloor \frac{(2^{k+1} - 1)\alpha + \phi - 1}{2} + 1 \right\rfloor \tag{10}$$

$$= \left\lfloor (2^k - 1)\alpha + \frac{\phi + \alpha + 1}{2} \right\rfloor. \tag{11}$$

□

3.1 Strongly balanced and mechanical trees are the same

In this section, we will show that if the density is irrational, a strongly balanced tree is a mechanical tree. This result can be expressed by the following proposition.

Proposition 3.4. *Let $\mathcal{A}$ be a infinite binary tree.*
(i) If $\mathcal{A}$ is mechanical then $\mathcal{A}$ is strongly balanced.
(ii) If $\mathcal{A}$ is strongly balanced and not rational, then $\mathcal{A}$ is a mechanical tree.
(iii) If $\mathcal{A}$ is strongly balanced and rational, $\mathcal{A}$ is ultimately mechanical.

When we say that a tree verifies a property P *ultimately*, that means that there exists a K such that each sub-trees of $\mathcal{A}$ which root if of height $k > K$, k verifies the property. In that particular case, we say that a tree is ultimately mechanical if all of its sub-trees of root of height superior to a certain K are mechanical trees.

Let us begin the proof by this first lemma.

Lemma 3.5. *If $\mathcal{A}$ is a mechanical tree, then $\mathcal{A}$ is strongly balanced.*

Proof. Let $\mathcal{A}$ be a mechanical tree and let $k, q \in \mathbb{N}$, $k \leq q$. Let look at the quantity:

$$x(n, n', k, q) = |h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n,q]})| - |h(\mathcal{A}_{[n',k]}) - h(\mathcal{A}_{[n',q]})|.$$

Using the well-known formula $x - x' - 1 < \lfloor x \rfloor - \lfloor x' \rfloor < x - x' + 1$, we can write the following equalities:

$$\begin{array}{rcl} x(n, n', k, q) & = & \lfloor (2^k - 1)\alpha + \phi_n \rfloor - \lfloor (2^q - 1)\alpha + \phi_n \rfloor - \lfloor (2^k - 1)\alpha + \phi_{n'} \rfloor + \lfloor (2^q - 1)\alpha + \phi_{n'} \rfloor, \\ & < & (2^k - 1)\alpha + \phi_n - (2^q - 1)\alpha - \phi_n - (2^k - 1)\alpha - \phi_{n'} + \lfloor (2^q - 1)\alpha + \phi_{n'} + 2, \\ & < & 2 \\ x(n, n', k, q) & > & -2. \end{array}$$

Moreover, as $x(n, n', k, q)$ is an integer: $|x(n, n', k, q)| \leq 1$ that proof the lemma. □

Let now proof the reciprocal of the lemma:

Lemma 3.6. *Let $\mathcal{A}$ be a strongly balanced tree with density α which can not be written as $\frac{p}{2^k - 2^q}$ $(p, k, q \in \mathbb{N})$.*
Then $\mathcal{A}$ is a mechanical tree.

Proof. (sketch) The first point is a rather direct consequence of the definition of mechanical trees. As for the second point, let τ be a real number and n a node. At least one of the two following properties is true:

- for all k: $h(\mathcal{A}_{[n,k]}) \leq \lfloor (2^k - 1)\alpha + \tau \rfloor$,
- for all k: $h(\mathcal{A}_{[n,k]}) \geq \lfloor (2^k - 1)\alpha + \tau \rfloor$.

To prove this, assume that it is not true. Then there exists k, q such that $h(\mathcal{A}_{[n,k]}) < \lfloor (2^k - 1)\alpha + \tau \rfloor$ and $h(\mathcal{A}_{[n,k+q]}) > \lfloor (2^{k+q} - 1)\alpha + \tau \rfloor$ (or the opposite).

In that case: $h(\mathcal{A}_{[n,k+q]}) - h(\mathcal{A}_{[n,k]}) \leq 2 + \lfloor (2^{k+q} - 1)\alpha + \phi \rfloor - \lfloor (2^k - 1)\alpha + \phi \rfloor > 1 + (2^{k+q} - 2^k)\alpha$ which violates the formula obtained in (2).

Let us define the number ϕ as follows:

$$\phi = \inf_{\tau} \left\{ \text{For all } k : h(\mathcal{A}_{[n,k]}) \leq \lfloor (2^k - 1)\alpha + \tau \rfloor \right\}.$$

Then we have for all k:

$$h(\mathcal{A}_{[n,k]}) \leq (2^k - 1)\alpha + \phi \leq h(\mathcal{A}_{[n,k]}) + 1.$$

If $\alpha \notin \{\frac{p}{2^k - 2^q}, p, k, q \in \mathbb{N}\}$, then $(2^k - 1)\alpha + \phi$ is an integer for at most one k_0. Depending on the value in k_0: if $h(\mathcal{A}_{[n,k_0]}) = (2^{k_0} - 1)\alpha + \phi$, then for all k: $h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi \rfloor$. Otherwise for all k: $h(\mathcal{A}_{[n,k]}) = \lceil (2^k - 1)\alpha + \phi - 1 \rceil$. ☐

Now what happens if α belongs to $\{\frac{p}{2^k - 2^q}, p, k, q \in \mathbb{N}\}$? The result is quite different and as we will see, since there it works *ultimately*. A counterexample is the tree which nodes are labeled by 1 except for one labeled by 0. This tree is clearly strongly balanced with density 1 but is not mechanical. An other one is the tree depicted in figure 8 which has density 1/3.

Lemma 3.7. *Let $\mathcal{A}$ be a strongly balanced tree.*
Then $\mathcal{A}$ is a ultimately mechanical.

Proof. As for the last assertion, the same as for the previous point holds until the fact that there is at most integer k_0 such that $(2_0^k - 1)\alpha + \phi \in \mathbb{N}$.

Now we assume that there exists k_0, k_1 such that $(2^{k_i} - 1)\alpha + \phi \in \mathbb{N}$ (otherwise we go to the end of the proof of (ii)).

A direct computation leads to $\alpha \notin \{\frac{p}{2^k - 2^q}, p, k, q \in \mathbb{N}\}$ and thus there exist p and $q_0 < q_1$ such that

$$\alpha = \frac{p}{2^{q_0} - 2^{q_1}}, \mathrm{GCD}(p, 2^{q_0} - 2^{q_1}) = 1.$$

According to the formula 2, this means that $h(\mathcal{A}_{[n,q_1]}) - h(\mathcal{A}_{[n,q_1]})$ equals $p-1$,p or $p+1$. But as the tree is strongly balanced, there can not be in the tree two sub-trees with values $p-1$ and $p+1$. Assume there is $q_1 - q_0$ other pairwise disjoint sub-trees with value $p+1$, then there would be two of them for which the roots will be at height h and $h + (q_1 - q_0) * n$ and the minimal sub-tree that contains this two sub-trees would violate the formula 2.

Thus ultimately, all sub-trees corresponding to $h(\mathcal{A}_{[n,q_1]}) - h(\mathcal{A}_{[n,q_1]})$ will take the value p. For the rest of the proof, we assume that our tree verifies this property.

Now, let us recall that for all k:

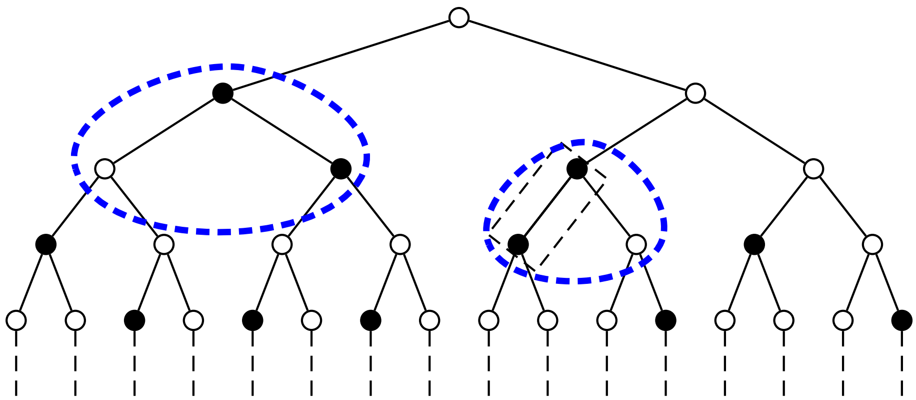
$$h(\mathcal{A}_{[n,k]}) \leq (2^k - 1)\alpha + \phi \leq h(\mathcal{A}_{[n,k]}) + 1.$$

If there are two integers k_0, k_1 such that $(2^{k_i} - 1)\alpha + \phi \in \mathbb{N}$ we have $\frac{2^{k_1} - 2^{k_0}}{2^{q_0} - 2^{q_1}} p \in \mathbb{N}$. Thus $\frac{2^{k_1} - 2^{k_0}}{2^{q_0} - 2^{q_1}} \in \mathbb{N}$. If we write $x = q_1 - q_0$ and $y = k_1 - k_0$, $y = px + r$ $(0 \leq r < x)$, we have: $\frac{2^{k_1} - 2^{k_0}}{2^{q_0} - 2^{q_1}} = 2^{k_0 - q_0} \frac{2^x - 1}{2^y - 1}$ which implies $\frac{2^x - 1}{2^y - 1} \in \mathbb{N}$, $\frac{2^x - 1}{2^y - 1} = 2^r \frac{2^{px} - 1}{2^x - 1} + \frac{2^r - 1}{2^x - 1} \in \mathbb{N}$, so $r = 0$ and $q_0 - q_1$ divides $k_1 - k_0$.

As for all n': $h(\mathcal{A}_{[n',q_1]}) - h(\mathcal{A}_{[n',q_0]}) = p$, if $h(\mathcal{A}_{[n,k_0]}) = (2^{k_0} - 1)\alpha + \phi$ then $h(\mathcal{A}_{[n,k_1]}) = (2^{k_1} - 1)\alpha + \phi$ and then for all k:

$$h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi \rfloor.$$

If $h(\mathcal{A}_{[n,k_0]}) = (2^{k_0} - 1)\alpha + \phi - 1$ then $h(\mathcal{A}_{[n,k_1]}) = (2^{k_1} - 1)\alpha + \phi - 1$ and then for all k: $h(\mathcal{A}_{[n,k]}) = \lceil (2^k - 1)\alpha + \phi \rceil$ ☐



This tree is rational with density 1/3, which means that all following sub-trees of height 2 that we can not see are trees with one 1 and two 0. One can verify on the picture that the beginning of this tree is strongly balanced and as it continues with density exactly 1/3, the whole tree is strongly balanced.
However this tree is ultimately mechanical but not mechanical because of the two sub-trees of height 2 surrounded by blue circles are not mechanical. If they were mechanical, we would have $\alpha = 1/3$ and a ϕ such that $\lfloor 3\alpha + \phi \rfloor = 2$ and then $\phi = 1$ which contradicts the definition. Moreover, if we look at the sub-tree of root 0, it is also not mechanical since we would have $\lfloor \alpha + \phi \rfloor = 0$ and $\lfloor 7\alpha + \phi \rfloor = 3$ which is impossible.

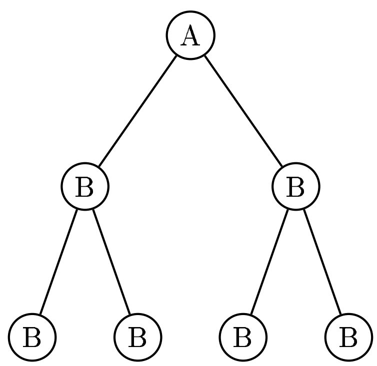


Figure 8: Example of a rational tree which is strongly balanced and ultimately mechanical but not mechanical.

Although we can characterize strongly balanced trees, it is still an open question to characterize balanced trees. The first answer is that there exist some trees that are balanced but not strongly balanced. We give an example with figure 9 where this picture is the beginning of a rational tree with density 3/7.

The previous example shows a rational tree which is balanced but not strongly balanced but as we have seen, rational trees admit more exceptions than irrational ones. However, there are multiple examples of trees that are irrational but not strongly balanced. Such a construction is more involved.

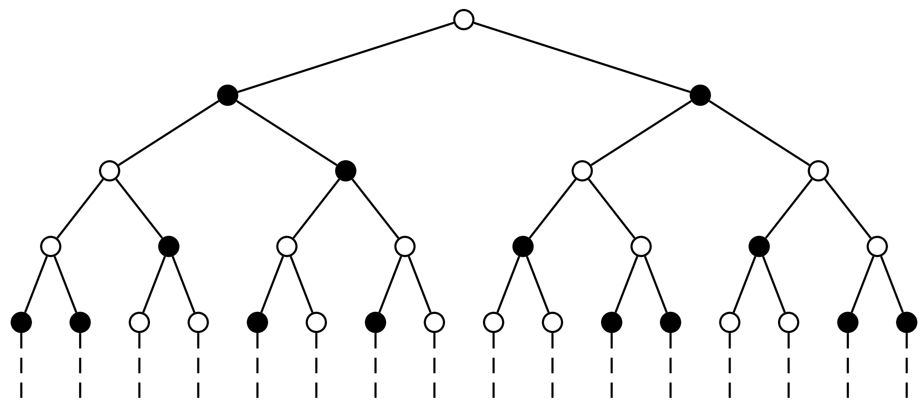
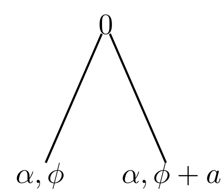


Figure 9: Beginning of a balanced tree of density 3/7 which is not strongly balanced (and thus non mechanical). This tree is rational: it has 3 sub-trees of height 3 (the 3 written on the figure). This tree is not strongly balanced since in the picture two non-routed subtrees with 0 and 2 ons resp. are highlighted. As this tree is rational, it is not ultimately strongly balanced either, since the highlighted motifs are repeated an infinite number of times.

We want to build a strongly balanced tree with density α irrational which root is labelled by 0 and has two children which are mechanical trees of density α and respective phases ϕ and $\phi + a$.



The two sub-trees with phases (ϕ) and ($\phi + a$) are balanced of density α, which means that for all $n > 0$, they satisfy the relation $\lfloor (2^k - 1)\alpha \rfloor \leq h(\mathcal{A}_{[n,k]}) \leq \lfloor (2^k - 1)\alpha \rfloor + 1$. Thus the whole tree is balanced *if and only if* all the sub-trees beginning at the root satisfies the relation:

$$\lfloor (2^k - 1)\alpha \rfloor \leq h(\mathcal{A}_{[0,k]}) \leq \lfloor (2^k - 1)\alpha \rfloor + 1. \tag{12}$$

As the two children are mechanical, we have

$$h(\mathcal{A}_{[0,k]}) = \lfloor (2^{k-1} - 1)\alpha + \phi \rfloor + \lfloor (2^{k-1} - 1)\alpha + \phi + a \rfloor$$

Let us now call $z = (2^k - 1)\alpha$, $y_1 = (2^{k-1} - 1)\alpha + \phi$, $y_2 = (2^{k-1} - 1)\alpha + \phi + a$. The relation 12 that we want to be satisfied can be written as

$$0 \leq \lfloor y_1 \rfloor + \lfloor y_2 \rfloor - \lfloor z \rfloor \leq 1$$

We have

$$\begin{aligned} z &= (2^k - 1)\alpha \\ &= 2(2^{k-1} - 1)\alpha + \alpha \\ &= y_1 + y_2 - 2\phi - a + \alpha. \end{aligned}$$

Let $x = \lfloor y_1 \rfloor$. We have the following equalities:

$$\begin{aligned} y_1 &= n + x \\ y_2 &= n + x + a \\ z &= 2n + 2x - 2\phi + \alpha. \end{aligned}$$

Depending on $x + a$ greater or less than 1, the relation is satisfied if

$$(x + a < 1 \text{ and } -1 \leq 2x - 2\phi + \alpha < 1) \text{ or } (x + a \geq 1 \text{ and } 0 \leq 2x - 2\phi + \alpha < 2).$$

From these inequalities, we get 4 relations:

$$\begin{aligned} 2(1 - a) - 2\phi + \alpha &\leq 1 \\ -1 &\leq -2\phi + \alpha \\ 2 - 2\phi + \alpha &\leq 2 \\ 0 &\leq 2(1 - a) - 2\phi + \alpha. \end{aligned}$$

From these relations, we get that the tree is balanced *if and only if*

$$\frac{\alpha}{2} \leq \phi \leq \frac{\alpha + 1}{2} \text{ and } \frac{\alpha + 1}{2} \leq \phi + a \leq \frac{\alpha + 2}{2}. \tag{13}$$

An example of such a tree is pictured figure 10.

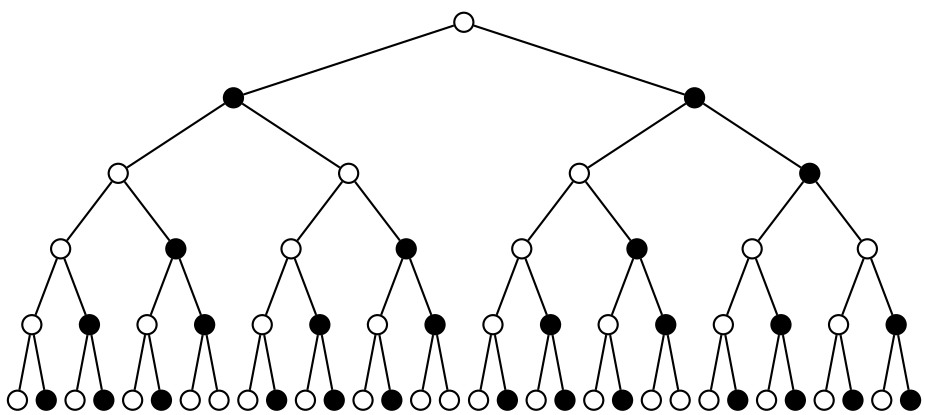


Figure 10: Example of a balanced tree with irrational density $3/7 + \epsilon$ which is not strongly balanced. It is not strongly balanced because of the two highlighted sub-trees . The children of the root have phases 0.6 and 0.8 so they satisfy Equation 13.

.

Example 1. *Another possible construction is a tree with density* $3/7 + \epsilon$ *whose root is labeled by* 0 *and its two children are mechanical words with respective* ϕ $4/7$ *and* $3/7$.

Building mechanical trees.

Using the constructive definition, it is very easy to build a prefix of a mechanical tree of finite length.

- A finite strongly balanced tree of height k can be built in linear time in the number of node $n = 2^k - 1$.
- The label of a node of height k can be computed in linear time in k.

Number of mechanical trees for $\phi = 0$. It seems hard to enumerate the number of strongly balanced trees.The numeration list starts with

$$2, 6, 20, 57, 158, 428, 1076, 2640, 6198, 14362, 32894, 73242, \ldots$$

As for mechanical trees with phase $\phi = 0$, if one calls C_k the number of such trees of size k, a tree corresponds bijectively to a sequence $(\lfloor (2^q - 1)\alpha \rfloor)$. Then the number of trees equals the number of times the line of slope α touches an integer number of the form $(2^q - 1, p)$.

$$C_k = C_{k-1} + 2^k - 2 - \sum_{\substack{q \mid k \\ k/q \text{ is prime}}} (2^q - 2).$$

To end this part, notice that two mechanical trees with the same density are very close. Also, two mechanical trees with different densities just have a finite number of factors in common.

Proposition 3.8. *Let $\mathcal{A}$ and $\mathcal{B}$ be two mechanical trees. Let $S(\mathcal{A})$ and $S(\mathcal{B})$ be the set of their respective sub-trees.*
(i) If the densities of $\mathcal{A}$ and $\mathcal{B}$ are the same, then $S(\mathcal{A}) = S(\mathcal{B})$.
(ii) If the densities are different, then $S(\mathcal{A}) \cap S(\mathcal{B})$ is finite.

3.2 Strongly balanced trees are Sturmian

Proposition 3.9. *Let $\mathcal{A}$ be a mechanical tree and $k \geq 0$.*
(i) There exists at most $k + 1$ sub-tree of height k
(ii) Moreover, if α is irrational, then the number of sub-trees is exactly $k + 1$.

Proof. Let $\mathcal{A}$ be a mechanical tree of density α and let $k \leq 0$. According to the proposition 3.2, the sub-tree $\mathcal{A}_{[n,k]}$ depends only on its phase ϕ_n. In fact, this sub-tree depends only on the values $\lfloor (2^i - 1)\alpha + \phi_n \rfloor$.

For all $i \leq k$ and $\phi \leq 1$, we can define functions $h_i(.)$

$$h_i(\phi) = \lfloor (2^i - 1)\alpha + \phi \rfloor.$$

These are increasing functions taking integer values and $h_i(1) - h_i(0) = 1$.

Thus the $k-$tuple $(h_1(\phi), \ldots, h_k(\phi))$ can take at most $k + 1$ values and there are at most $k + 1$ sub-trees.

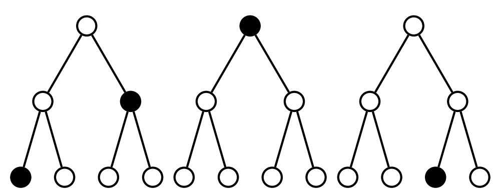
Moreover if α is irrational, the tree has an irrational density and then it is not rational so there are at least $k + 1$ trees of height k which means that the tree has exactly $k + 1$ sub-trees of height k. ☐

As shown in the previous section, if a tree has k factors of size k, it is rational. Thus strongly balanced trees of irrational density are trees with minimal complexity. We know that in the case of words, aperiodic balanced word are exactly words with complexity $n+1$ however this is not the case here. [2] gives some examples of planar Sturmian trees, although some of theses examples do not work in the non-planar case, many do.

- *Uniform trees*: considering a word w, define the uniform tree for w as the tree where a node with height k is labeled by w_k. If w is Sturmian, this tree is also Sturmian (*i.e.* with complexity $n+1$),
- *Left branch trees*: considering a word w. Here is how the definition works in the planar case: the label of a node n is w_k where k is the number of time with have to go left on a path from the root. Even if we consider the non-planar version of this tree, it is clearly Sturmian *iff* w is Sturmian.
- *Dyck tree*: recall that in the planar case, a node can be represented by a word on $\{0, 1\}$ representing the path from the root. The Dyck tree is the tree where the label of a node is 1 *iff* its representing word belongs to the Dyck language. The non-planar version of this tree is also Sturmian.

On the over hand, there exits balanced trees with a complexity less than $n+1$ which are not strongly balanced, see the tree shown in the following example.

Example 2. *Consider a tree which factors are:*



This tree is balanced, has complexity less than $n+1$, and it is not strongly balanced.

3.3 Trees with complexity $n+1$

For Sturmian words, balance and minimal complexity are almost equivalent properties. For trees, this is not the case since there exist trees which have complexity $n+1$ which are not balanced.

Example 3. *Consider any Sturmian word m and construct a tree such that the i-th level nodes all have label m_i. This tree is not balanced but has complexity $n+1$.*

Example 4. *Other trees such as the Left Branch Tree or the Dyck tree are not balanced and do not have a pseudo density, with a complexity $n+1$.*

4 Summary and perspectives

Figure 11 gives an synthetic view of the results presented in this paper under the form of set intersections.

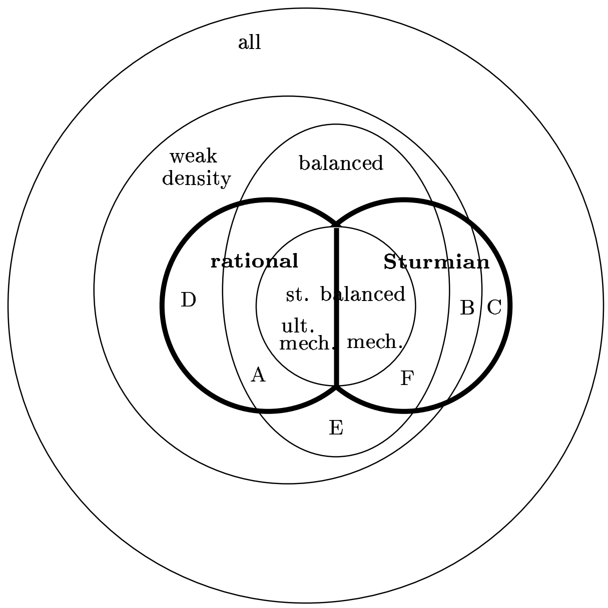


Figure 11: Classification of all non-planar trees. A tree belonging to set A is given in example 2. A tree in set B is given in example 3. A tree in set C is given in example 4. A tree in set D is the rational tree with only two factors 000, 111 and an initial root factor 101. A tree in set E is given in Figure 10 and F can also be shown to be non empty (a tree in F is given in example 1).

Several questions remain, in particular concerning random generation of balanced trees, links with continuous fractions and applications to optimization problems as a generalization of what has been done for sturmian words in [1].

A balanced tree has the strong property that there is a density α of nodes labeled by 1 in each sub-tree.

Let $g : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex function. For each node n and each height $k > 0$, we define a cost $C_{[n,k]}$:

$$C_{[n,k]} = g(d(\mathcal{A}_{[n,k]})).$$

Let us assume that g has a minimum in α.

We can define a cost of order k of the tree by

$$C_k = \limsup_{l \to \infty} \frac{\sum_{n \in \mathcal{A}_{[0,l]}} C_{[n,k]}}{2^l - 1}.$$

For each k, this cost is minimized when the number of 1 in a tree of height k is between $\lfloor \alpha(2^k - 1) \rfloor$ and $\lceil \alpha(2^k - 1) \rceil$. That means that a balanced tree will minimize any increasing function of all C_k. This has potential applications in optimization problem in distributed systems with a binary causal structure and would generalize the results in [1].

References

[1] E. Altman, B. Gaujal, and A. Hordijk. *Discrete-Event Control of Stochastic Networks: Multimodularity and Regularity.* Number 1829 in LNM. Springer-Verlag, 2003.

[2] J. Berstel, L. Boasson, O. Carton, and I. Fagnot. A first investigation of sturmian trees. In *STACS'2007*, volume 4393 of *LNCS*, 2007.

[3] E. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *R.C. Mat.Palermo, t.27, 1909 np. 247-270*, 1909.

[4] M. Lothaire. *Algebraic Combinatorics on Words*, chapter Sturmian Words. Springer, 2002.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399